

Multiscale processing of loss of metal: a machine learning approach

G De Masi¹, M Gentile¹, R Vichi¹, R Bruschi¹ and G Gabetta²

¹Advanced Engineering (ADVEN) Dept. Saipem, Fano, Italy

²SIASV Dept., eni, San Donato Milanese, Italy

Email: giulia.demasi@saipem.com

Abstract. Corrosion is one of the principal causes of degradation to failure of marine structures. In practice, localized corrosion is the most dangerous mode of attack and can result in serious failures, in particular in marine flowlines and inter-field lines, arousing serious concerns relatively to environmental impact. The progress in time of internal corrosion, the location along the route and across the pipe section, the development pattern and the depth of the loss of metal are a very complex issue: the most important factors are products characteristics, transport conditions over the operating lifespan, process fluid-dynamics, and pipeline geometrical configuration. Understanding which factors among them play the most important role is a key step to develop a model able to predict with enough accuracy the sections more exposed to risk of failure. Some factors play a crucial role at certain spatial scales while other factors at other scales. The Mutual Information Theory, intimately related to the concept of Shannon Entropy in Information theory, has been applied to detect the most important variables at each scale. Finally, the variables emerged from this analysis at each scale have been integrated in a predicting data driven model sensibly improving its performance.

1. Introduction

CO₂ corrosion, or the so called “sweet corrosion”, is a major concern in the application of carbon and low alloy steels, which are still the principal construction materials, offering economy, availability and strength. External sea water corrosion is prevented by the use of an external coating (passive protection) and by cathodic protection (active protection). While for long transportation pipelines the gas is normally dry and the internal corrosion is not an issue, for flowlines and inter-field lines transporting untreated fluids, internal corrosion plays a crucial role for structure integrity. This is a growing and challenging problem for Oil and Gas industry, since the age of plants and components is worldwide increasing. It is necessary to prevent failures of all these components.

Internal corrosion shows a very complex phenomenology due to interaction of different mechanisms playing a role at different spatial scales. Water and electrochemistry, protective scales, flow velocity, steel composition and localized bacteria attacks are the most relevant.

While it is reasonably easy to understand a corrosion event “retrospectively” with failure analysis methods, a large degree of uncertainty is associated with the attempt of quantifying a prediction for the future evolution of damage.

Relevant physics analysis is multi-disciplinary and includes different tasks: analysis of products, transport conditions over the operating lifespan, process fluid-dynamics and pipeline configuration on the real seabed [1].

An effective corrosion modelling methodology would be based on fundamental laws and first principles since analyses could be performed over a wide range of conditions with some degree of confidence. But corrosion mechanisms are not yet understood well enough to develop purely first-



principles mechanistic models, although this is the final goal of research on corrosion. For this reason semi-empirical and empirical models have been studied. Unfortunately, due to its complexity, the localized corrosion process is hardly reproduced even by semi-empirical models proposed in literature [2, 3]. Furthermore, differently from other publications focusing on prediction of generalized corrosion, the scope of the present paper is to identify along the same pipeline where the sections with high risk of corrosion are located.

Field data, nowadays increasingly available, e.g. from internal line inspections (ILI) by intelligent pigs, allows the development of data-driven fully empirical models that can be useful to process corrosion data [4]. Given that the uncertainties related to measurements of remaining wall thickness in an intelligent pig are very low, this allows to accurate estimate volume loss and corrosion rate along a specific pipeline. Data driven model can also shed light on which factors are more relevant for the localized corrosion process, with the final aim to develop an analytic model.

This paper offers a methodology to identify the crucial factors playing the most important role in the corrosion process under study. The methodology is based on Mutual Information Theory strictly related to the Shannon Entropy theory[5]. The mutual information (MI) of two random variables is a measure of the mutual dependence between the information content of two variables, estimated by their Entropies $H(X)$, $H(Y)$ like in figure 1.

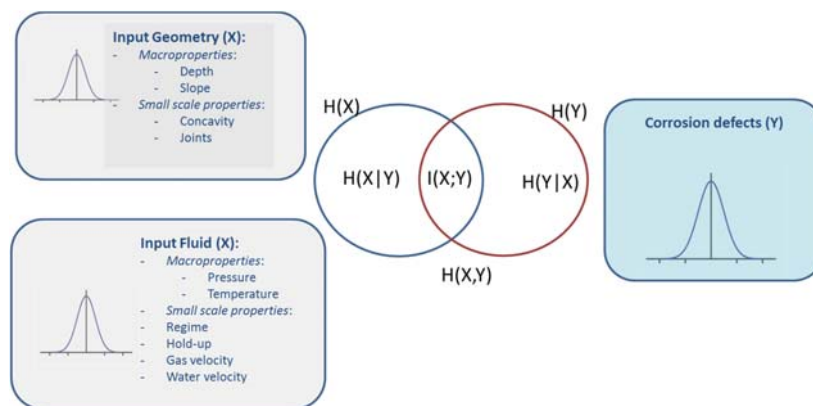


Figure 1. Scheme of the Mutual Information Analysis.

More specifically, it quantifies the "amount of information" obtained about one random variable, through the other random variable. In our particular case it is used to understand how the observed corrosion pattern depends on each possible external condition (from material property, to fluid transport conditions, to pipeline geometry, to microbiology). Analyses have been performed at different spatial scales, being the phenomenon partially global on the whole length of the pipe (generalized corrosion), partially medium scale (corrosion due to increase/decrease of fluid-velocity or passage of slugs), partially very local scale (for instance, stagnant water regions).

Finally the most important variables are fed into a predictive model. Among several expert systems and data driven modelling, in this paper artificial neural networks (ANN) have been chosen [6]. While application of ANNs has already been demonstrated to be very useful for internal corrosion prediction[7, 8], the authors apply the MI method to improve a predictive model based on a set of cascading ANNs which has been demonstrated to be particularly effective to identify the sections with higher corrosion up to a detailed scale of 10m [9].

The MI analysis strongly improves the prediction performance through the identification of the most important variables at each scale, compared with past papers where all the contribution were considered equivalent[9].

2. Methodology

2.1. Product characterization and historical data collection

First step is collection and organization of all existing, relevant, essential, historic and current operating data about the pipeline segments and/or regions relevant to corrosion distribution.

From design and construction records, operating and maintenance histories, corrosion survey records, gas and liquid analysis reports, and inspection reports from prior integrity evaluations or maintenance actions.

2.2. Multiphase flow regimes

Once identified a representative scenario of fluid-dynamic boundary conditions, multiphase flow modelling is done using OLGA Multiphase Flow Simulator software. As the operating conditions vary along the route, this program determines temperature profile along the pipeline, pressure profile, velocity profiles of each phase, hold-ups and flow regimes, given boundary pressure, temperature values and flow composition.

2.3. Geometric profile

The choice of a significant profile and consequently inclination and concavity to set as input for ANN is a very crucial issue. Generally, a profile, i.e. KP vs. Elevation, provided by an ILI inspection is quite accurate. In order to obtain significant variations of the geometrical variables, a filtering of the profile has been performed. The filter is chosen such that the final profile is representative of main spatial variations, without having too much detail.

2.4. Mutual Information Theory

As evident, one crucial point in the corrosion dataset processing is to identify which are the most relevant variables for the predictive model. In order to evaluate the dependence between the observed corrosion and the other variables, the Mutual Information (MI) analysis has been implemented. MI or (formerly) trans-information of two random variables X and Y is a measure of the variables mutual dependence (Fig.1). Not limited to real-valued random variables like the correlation coefficient, MI is more general and determines how similar the joint distribution $p(X,Y)$ is to the products of factored marginal distribution $p(X)p(Y)$ and it is given by

$$MI(X,Y) = \sum_i p(x_i, y_i) \log \frac{p(x_i, y_i)}{p(x_i)p(y_i)} \quad (1)$$

It can be demonstrated that the relation of MI with Information Entropy of single variables $H(X)$, $H(Y)$ and their cross Entropy $H(X,Y)$ is given by:

$$MI(X,Y) = H(X) + H(Y) - H(X,Y) \quad (2)$$

In contrast to the linear correlation coefficient, it is sensitive also to dependencies which do not manifest themselves in the covariance. In the present study, the variable Y is the corrosion volume loss while the variable X comes from four different categories:

- Hydrocarbon characteristics (CO_2 partial pressure)
- Geometrical pipeline characteristics (elevation, inclination, concavity)
- Fluid dynamic multiphase variables (flow regime, hold-up, pressure, gas flow, total flow, liquid velocity, gas velocity)
- Deterministic models (de Waard and NORSOK)

2.5. Predictive model

In order to represent multi-scale phenomenology and to obtain prediction of regions more exposed to corrosion also at fine scale (for our purposes 10m), a set of cascading ANNs, each one representing the phenomenon at a different spatial scale (1000m, 100m, 10m) has been designed [9].

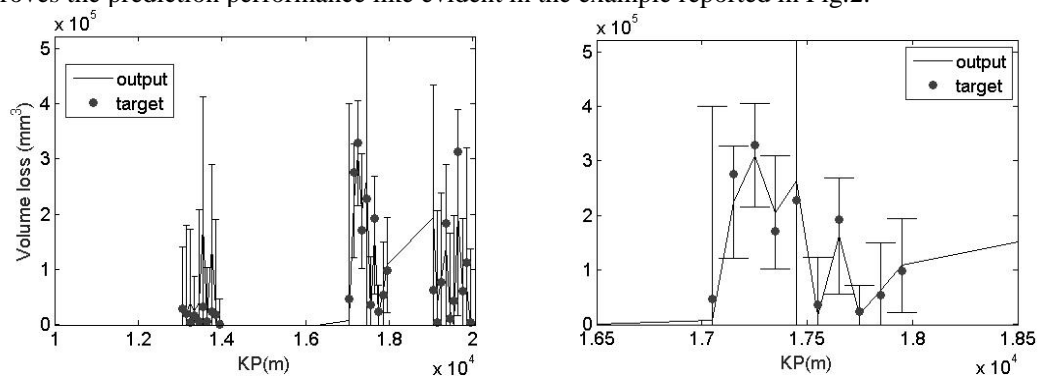
3. Results and Conclusions

The MI analysis has been performed at different spatial scales (1000m, 100m and 10m). The results are reported in table 1.

Table1. Results of Mutual Information Analysis. Only variables with Mutual Information over 0.5 are reported.

1000m		100m		10m	
Elevation	0.75	Pressure	0.63	Hold-up	0.84
KP	0.59	Elevation	0.63	Concavity	0.64
Liquid Velocity	0.57	Gas Velocity	0.63	Liquid Velocity	0.60
Pressure	0.57	Hold-up	0.59	GasVelocity	0.50
Gas Flow	0.57	Concavity	0.53		
Hold-Up	0.57	Gas Flow	0.53		
Regime	0.53				

The variables well represent the dynamics at different spatial scale. Feeding the variables selected by MI analysis on the predictive hierarchical model, the final prediction at fine scale (10m) strongly improves the prediction performance like evident in the example reported in Fig.2.

**Figure 2.** Example of output of the model.

Despite the fact that the observed corrosion pattern is very complex, the integration of multiscale MI analysis and the predictive model seems to capture main evidences of local corrosion, providing a promising methodology for crucial assessment. Analysis and predictions can be further improved in the future considering larger datasets (with several pipeline cases and different flow conditions).

References

- [1] Bruschi R, Cherubini P, Vichi R, Gentile M 2013 *Integrity Management On Pipeline Affected by Corrosion* IBP1471_13 Rio Pipeline Conference & Exposition 2013 Rio de Janeiro (Brazil)
- [2] De Waard C and Lotz U 1993 *NACE Corrosion 1993 Proc.* (Houston) Paper N°69
- [3] *NORSOK Standard N°M-506 Rev1 Norwegian Technology Standards Institution* (<http://www.wntsnor/norsok>)
- [4] Li X, Zhang D, Liu Z, Li Z, Du C and Dong C 2015 *Nature* **527** 441-442
- [5] Gray R M 2008 *Entropy and Information Theory* (Springer)
- [6] Haykin S 1999 *Neural Networks A comprehensive foundation* (Pearson Prentice Hall)
- [7] Gabetta G, De Masi G, Gentile M, Vichi R and Scapin M 2014 *ADIPEC Proc. SPE* **171919**
- [8] De Masi G, Vichi R, Gentile M, Bruschi R and Gabetta G 2015 *Oceans 2015 Proc.* (Genova)
- [9] De Masi G, Vichi R, Gentile M, Bruschi R and Gabetta G 2016 *Compit 2016 Proc.* (Lecce)